ISSN 0703-8763

Edmonton Area Series Report No. 60

ITEM ORDER EFFECTS ON THE MEASUREMENT OF ABORTION ATTITUDES: A QUASI-EXPERIMENTAL DESIGN

by

Elisabeth TenVergert
Michael W. Gillespie
Johannes Kingma

University of Alberta

The 1984 Edmonton Area Study (EAS) and the 1987 All Alberta Study (AAS) were administered and supported by the Population Research Laboratory and the Department of Sociology at the University of Alberta.

July 1988

Abstract

A vexing problem in measuring change in attitudes is separating the change due to the sequence of the questions in the surveys from real change. The present study investigates empirically the extent to which the responses to a set of attitude items are effected by the order of presentation. Another concern of the study was to determine the effects of the item order manipulations on the scalability of the items.

We addressed these issues by analyzing the six abortion items that appeared in the 1984 Edmonton Area Survey and the 1987 All Alberta Study. We found that questionnaire context effected significantly the level of endorsement of the manipulated items.

Furthermore, we used the Mokken scaling method to investigate whether the item manipulations would effect both the scalability and dimensionality of the six items. It was shown, that notwithstanding the strong item order effects, the unidimensionality of the six abortion items remained stable in both surveys. We conclude by describing how this scale can be used to measure the change in attitudes toward abortion across the two periods.

Item Order Effects on the Measurement of Abortion Attitudes

The effect of item order in survey research makes the measurement of social change difficult. The problem is separating the change due to the sequence of the items in the surveys from real change. The sequence of the items is often not held constant from one survey to another because certain items that appear in surveys conducted during the early periods are likely to be dropped as they become rendered obsolete by social change, while other items will apear in later surveys in order to capture new circumstances that relate to the attitude measured.

The impact of the order of items on the responses to questions has been amply investigated (Sigelman, 1981; Schuman et al., 1981; McFarland, 1981; Kraut et al., 1975; Duncan and Schuman, 1980; Bradburn and Sudman, 1979). Although the findings of these studies show mixed results, enough evidence has turned up that placing an item in different positions in various questionnaires may effect the responses to the questions. However, a general theory that will tell researchers when such order effect takes place or how to avoid them is still lacking (Bradburn, 1979).

One common recommendation is that questions which ask for a general evaluation on a particular issue should precede questions about its more specific aspects (Kornhauser and Sheatsley, 1976). For example, Schuman et al. (1981) found that responses to a general question about abortion were significantly greater when asked first than when asked after a more specific question on abortion in the case of a defect in the unborn child. They call this a part-whole effect because agreement to the general abortion item ordinarily implies agreement to the more specific defective-child item, but the reverse is not true.

One concern of the present study is to investigate the extent to which the

responses to questions on abortion are affected by the order of presentation. To this end, we propose to analyze data from the 1984 Edmonton Area Studies (EAS) and the 1987 All Alberta Study (AAS). In both surveys, respondents were asked whether they approve of legal abortion under the following conditions: (a) a strong chance of a serious defect in the baby (DEFECT), (b) the woman is married and does not want any more children (NOMORE), (c) the woman's own health is seriously endangered by the pregnancy (HEALTH), (d) the family has a very low income and cannot afford any more children (POOR), (e) the woman became pregnant as a result of rape (RAPE), and (f) the woman is not married and does not want to marry the man (SINGLE).) The sequence in which the items DEFECT and HEALTH were offered the respondents was manipulated in the questionnaire of 1987, i.e., the NOMORE item was treated as a "constant" item whereas the more specific items HEALTH and DEFECT were reversed in ordering.

Another purpose of the study was to investigate the effects of this manipulation on the dimensionality of the six abortion items for both years. Previous studies suggest that changes in question order may alter the frequency distributions of the items as well as the pattern of relationship between the responses given to the items (Duncan and Schuman, 1977; Hayes, 1964). As a consequence, comparisons of attitudes across different time periods may be limited since change in attitudes could simply reflect differences due to the effects of item order. In order to measure "real" change, however, we have to keep separated change in the general attitude measured by the set of items from change in the responses to specific items. For this reason, it is important to construct a unidimensional scale for the six abortion items which is equivalent across the two periods.

Unfortunately, previous research failed to resolve the issue of whether the different conditions under which abortion might be used constitute a

unidimensional set of items. Some studies treat the items as measuring a single dimension (Combs and Welch, 1982; Ebaugh & Haney, 1980; Finlay, 1985; Granberg, 1978; Hall & Ferree, 1986; Jones and Westhoff, 1973; Hartnagel et al., 1985), and others as measuring two unidimensional scales (Arney and Trescher , 1976; Barnett and Harris, 1982; Granberg and Granberg, 1980). These studies use the DEFECT, HEALTH, and RAPE items as measure of attitudes toward abortion for "medical" or "hard" reasons and the NOMORE, POOR, and SINGLE items as a measure of attitudes toward abortion for "social" or "soft" reasons.

The major empirical difference between the two sets of items is that the percentage of respondents who approve abortion for medical reasons is substantially greater than the percentage who approve abortion for social reasons. Put another way, the medical items are less "difficult" than the social items. But whether this difference justifies the separation of the six items into two different measures is another question. To be sure, factor analysis of the items separates the medical and social items into two factors when one adopts the convention of extracting as many factors as there are eigenvalues greater than one (Arney and Trescher, 1976; Granberg, 1978; SPSS INC., 1983:656-659). However, one could also interpret these results as an artifact of the differences in the item difficulty (Gillespie et al., 1987a, 1987b, in press; TenVergert, 1988). It is well known in the psychometric literature that factor analysis of dichotomous items that are homogenous in content but differ in the degree of difficulty typically yield two or more difficulty factors (Nunally, 1978; Ten Berge, 1972).

It is important to attempt to resolve the question of unidimensionality. When two subscales are produced by summing the hard and soft items separately, the researcher will likely find that the two scales will correlate differently

with variables believed to be related to abortion attitudes. If the two subscales measure different dimensions, such differences raise substantive questions that warrant empirical research. Why, for example, would attitudes toward abortion for medical reasons correlate less with schooling than attitudes toward abortion for social reasons? On the other hand, if the two subscales measure a single dimension, differences in the correlations would represent an artifact of differences in the difficulty of the items that comprise the two scales. One who pursues substantive explanations for these differences would be chasing a chimera.

Unfortunately, the studies which treat the measure as unidimensional offer no compelling evidence for their interpretation. First, some studies implicitly assume unidimensionality by simply summing the positive responses to each of the six items without presenting any evidence of scalability (Combs and Welch, 1982; Finlay, 1985; Hall and Ferree, 1986). Second, other studies subject these items to a Guttman (1950) scale analysis (Hartnagel et al., 1985; Arney and Trescher, 1976; Ebaugh and Haney, 1980; Granberg, 1978; Jones and Westhoff, 1973). Because of the clear difference in the difficulty of these items, we regard this approach as more appropiate than either factor analysis or the uncritical construction of a Likert scale and, therefore, worthy of additional discussion.

A well-known feature of Guttman's model is its specification that a positive response to an item of a given level of difficulty implies a positive response to all easier items. According to Guttman's model, all respondents who approve of abortion in the case of a single woman will approve of abortion when the woman's health is endangered. Because of this feature the individual's pattern of responses to the items can be reproduced from the total number of positive responses when the Guttman model holds (Guttman,

1950). A problem with Guttman's model, however, is that it is a deterministic one and, thus, makes no provision for measurement error. Because error is always present in measures of social attitudes, perfect Guttman scales occur seldom, if ever (Guttman 1950; Mokken, 1971). As a consequence, Guttman (1950) developed a reproducibility (REP) coefficient, based on the proportion of respondents whose response pattern falls into one of the scale types, in order to judge the fit between the data and Guttman's model.

All the studies of abortion attitudes that use Guttman's method report REP coefficients greater than .90. However, Jones and Westhoff (1973; p.14) point out that the improvement over minimal marginal reproducibility is not good. Jones and Westhoff's qualification is noteworty since it highlights a well-known problem with the use of Guttman's REP coefficient to assess the unidimensionality of a set of items: One can obtain very high values even though the items lack unidimensionality (Clogg and Sawyer, 1981; Guttman, 1950; Mokken, 1971; Nunally, 1978). For this reason, we fail to find the high REP coefficients for the abortion items compelling evidence of their unidimensionality.

From our point of view, however, the problem with Guttman scaling is not the use of reproducibility coefficients per se, for under the right circumstances, more appropiate coefficients than Guttman's REP can provide useful information. In order to do so, these coefficients need to be grounded in a more appropiate model of item responses: a probabilistic one rather than a deterministic one.

The Mokken Method of Scale Analysis

In light of the above discussion, the Mokken method of scale analysis should provide a useful method for assessing the unidimensionality of the abortion items (Mokken, 1971; Kingma and Reuvekamp, 1986a, 1986b; Gillespie et

al. 1987a; 1987b, Gillespie et al., in press). As a latent trait method, the Mokken model represents a nonparametric stochastic extension of Guttman's method for the scale analysis dichotomized items within a single population (Mokken, 1971). It is nonparametric in the sense that it imposes no fixed form on the function of the relationship that links the items to the latent variable (the trace line). Instead, it requires only that the relationship be monotonic and that the trace lines which describes this relationship do not intersect. Conseqently, items that meet the criteria of the Mokken model constitute an ordinal, rather than an interval or ratio scale.

Items that fit the Mokken model must satisfy the following four assumptions (see Appendix A). First, the items must be unidimensional, or dominated by the latent variable of interest. Second, local independence of item responses must hold for a fixed value on the latent variable. Third, the item trace lines must be nondecreasing, which implies a monotone relation between the probability of a positive response to the item and a respondent's value on the latent trait, that is, the items must meet the assumption of monotone homogeneity. Fourth, monotonicity in item difficulties must hold, i.e., there must be an identical ordering of probabilities of positive responses on a set of items for all values on the latent trait. (See Appendix A).

The latter two assumptions taken together is called the requirement of double monotony (Mokken, 1971). In order to analyze whether a set of items meet the four assumptions of the Mokken model, the test of the assumption of double monotony is the most crucial one, since this test is based on the assumption of local independence which underlies most latent trait and latent class models.

According to the assumption of double monotony, the probability of positive

response increases monotonically as the latent variable increases, and the ordering of the items by the level of item difficulty is invariant across the values of the latent variable. To test the requirement of monotone homogeneity, Mokken (1971) developed three related coefficients of scalability. The first, H_{ij}, measures the homogeneity or assocation between each pair of items i and j. The second, H_i, measures the homogeneity of a particular item with respect to all other items and is obtained by aggregating across the coefficients of the relevant item pairs. The third, H, measures the homogeneity of the scale as a whole and is obtained by aggregating across the individual coefficients of the individual items. (See Appendix A). Using a variety of criteria, Mokken (1971 pp. 184-185) judges these coefficients to be superior to Guttman's REP-coefficient and other coefficients of unidimensionality.

The researcher can test the monotonicity in the item difficulties by analyzing two matrices: the P- P_0 matrices. The cell entries for the P-matrix are the proportion of respondents who give positive responses to the item pair that defines the cell. The cell entries for the P_0- matrix contain the proportion of respondents who give negative responses to the two items. When the assumption of double monotony holds, the cell entries of the P-matrix will decrease with increasing item difficulty, while the cell entries of the P_0matrix will increase.

Because of the features described above, the Mokken method of scale analysis should help us study the dimensionality of the abortion items. In addition, we can use the Mokken method to see whether the order in which the items are presented effects the distribution of approval of the abortion items as well as the results of the Mokken method of scale analysis.

The Data

The data we analyze in this paper come from the 1984 Edmonton Area Study (EAS) and the 1987 All Alberta Study (AAS) both administered by the Population Research Laboratory and the Department of Sociology, University of Alberta. The 1984 EAS sample universe consists of the corporate boundaries of the City of Edmonton, whereas the 1987 AAS include all of Alberta. For purposes of comparability the present paper analyses only the Edmonton responses of the Alberta population included in the AAS 1987.

In both the 1984 and 1987 surveys, a simple random sample of households was selected from a computerized list of addresses compiled by the City of Edmonton from the 1983 and 1986 enumeration, respectively. Within the households respondents were selected from household members over eighteen years of age and were chosen on a sex quota basis, i.e., interviewers were instructed to obtain an equal number of male and female respondents within their allotment of interviews. In both surveys data were collected by face-to-face interviewing.

In the 1984 EAS a sample was drawn of 560 addresses and the final sample size was 452 giving a 80.7% response rate. In the 1987 AAS 620 addresses were selected for personal interviewing and the final sample size was 454 which represents a 76% response rate. Both surveys share a common set of items such as the six questions about the respondent's approval or disapproval of legal abortion [for further details on the 1984 EAS and the 1987 AAS sampling design see Kinzel (1984, 1987)].

We limit our analysis to just those respondents who gave valid responses (Yes or No) to all six items. This restriction reduced the number of cases analyzed to 395 in the 1984 EAS and 375 in the 1987 AAS.

Results

Crosstabulations of the two items DEFECT and NOMORE are presented in Table 1, separately by year, so that both the marginal differences and the intercorrelations can be compared for the two orderings. Examination of the marginals reveals that the NOMORE item received more support (by 12.5%) when asked before the DEFECT item and this difference is statistically significant (χ^2=11.42, df=1, p<.01). Table 1 also shows that the DEFECT item was affected by the order it was presented to the respondents. Placing the item after the items HEALTH and NOMORE as we did in the 1987 AAS, yielded an increase in approval by 8% (χ^2=12.34, df=1, p<.01).

Furthermore, Table 2 shows that the item HEALTH is unaffected by its placement (χ^2=2.28, df=1, n.s.), i.e., the proportion of respondents who approve legal abortion for HEALTH reasons is essentially unaffected by question order. Note further that in 198/ the marginal distributions of the HEALTH versus BOMORE items and the DEFECT versus NOMORE items are exactly the same.Except for the HEALTH item, the responses to the other two items NOMORE and DEFECT were affected by the context in which they appear. This result raises the question of whether these differences due to questionnaire context influence the measurement of the dimensionality of the abortion items.

To answer this question we turn to the second research question raised above by investigating whether the item rearrangements we made in the 1987 AAS effected the scalability of the six abortion items.

The proportion of respondents who endorse abortion under the six conditions is given for the two surveys in Table 3. Note how the differences in these proportions lend themselves to grouping the items into two sets: medical and social reasons. Also note that the proportion who approve abortion increases in 1987 and this is, in particular, the case for the items NOMORE and DEFECT. Note also that these two items were significantly affected by the

Table 1: The Effect of Question Order on Responses to Abortion Items.

Order 1984:DEFECT/NOMORE

		DEFECT: Yes	DEFECT: No	
NOMORE	Yes	42.0	1.3	43.3
	No	43.6	13.1	56.7
		85.6	14.4	100%
				(395)

Order 1987: NOMORE/DEFECT

		DEFECT: Yes	DEFECT: No	
NOMORE	Yes	55.5	.3	55.8
	No	38.1	6.1	44.2
		93.6	6.4	100%
				(375)

DEFECT marginals by order: χ^2=12.34, df=1, p<.01

NOMORE marginals by order: χ^2=11.42, df=1, p<.01

Table 2: The Effect of Question Order on Responses to Abortion Items.

Order 1984: NOMORE/HEALTH

		HEALTH: Yes	HEALTH: No	
NOMORE	Yes	42.7	.5	43.3
	No	47.7	9.1	56.7
		90.4	9.6	100%
				(395)

Order 1987: HEALTH/NOMORE

		HEALTH: Yes	HEALTH: No	
NOMORE	Yes	55.5	.3	55.8
	No	38.1	6.1	44.2
		93.6	6.4	100%
				(375)

HEALTH marginals by order: χ^2= 2.38, df=1, n.s.

NOMORE marginals by order: χ^2=11.42, df=1, p<.01

Table 3: Distribution of Approval of the Abortion Items: 1984, 1987.

Item	Difficulty Level*	
	1984	1987
SINGLE	.40	.41
POOR	.42	.44
NOMORE	.43	.56
RAPE	.83	.85
DEFECT	.86	.94
HEALTH	.90	.95

*Proportion of respondents who endorse the abortion item.

questionnaire context in which they appeared. Furthermore, 1984 EAS and 1987 AAS produce the same item orderings. The two questions these results raise is, first, whether the social and medical items represent separate subscales and, second, whether the drastically increase in approval of the DEFECT and NOMORE items affects the dimensionality of the items.

When we turn to statistics that pertain to the unidimensionality of the six items, we see first that the pattern of cell entries in the $\underline{P}$ and $\underline{P}_Q$-matrices support the assumption of double monotony. These matrices for both surveys are given in Tables 1B and 2B in Appendix B. Since none of the comparisons among the cell entries in the matrices yield differences opposite from the ones predicted by the Mokken model and the item difficulties (the proportion of respondents who endorse abortion), we regard these results as an indication that the six abortion items measure a single dimension.

Second, when we look at the individual item homogeneities in Table 4, we see that this index is quite high for both years. They range from .78 for SINGLE, NOMORE, and DEFECT in 1984 to .96 for DEFECT and HEALTH in 1987. Finally, the $\underline{H}$-coefficient of homogeneity for the entire set of items is .80 and .89 for the 1984, and 1987 periods, respectively. When we test these coefficients statistically, we find, first, that all are significantly greater than zero. Thus, we can reject the null hypothesis that the six items are independent. In addition, for both years the magnitude of this coefficient is significantly greater than .50, which Mokken (1971, p.185) uses to distinguish "strong" from "medium" Mokken scales. Thus, these results answer the second question raised above by indicating that the six abortion items form a unidimensional scale in both years.

In addition, the strong scales that emerge from separate Mokken analyses for the 1984 EAS and 1987 AAS samples indicate that the dimensionality of the

Table 4: The H_i Values for the Abortion Items: 1984 and 1987.

Item	H_i	
	1984	1987
SINGLE	.78	.88
POOR	.80	.87
NOMORE	.78	.87
RAPE	.80	.91
DEFECT	.78	.96
HEALTH	.87	.96

six abortion items is reasonably robust against the order in which the items are presented to the respondents. Thus, we could measure the "real" change in attitudes toward abortion for the years 1984, and 1987 although the order of item presentation differed in both surveys.

Discussion

The focus of this paper has been first, to examine the efects of questionnaire on the responses to specific abortion items, and second to determine how the change in level of endorsement of these items affect the unidimensionality of the whole set of items.

The findings suggest that there may be remarkable context effect on the item NOMORE , i.e., the NOMORE item received substantially more support when asked after the item HEALTH (AAS 1987) than when asked after DEFECT (EAS 1984). The latter item was also affected by the order of presentation: agreement with the DEFECT item was 8% lower when it comes first rather then after the NOMORE item. Furthermore, the HEALTH item was insensitive to questionnaire context.

Schuman et al. (1983) observed similar context effects for thethe abortion items. They found that approval of the general item NOMORE was considerably higher when asked first than when asked after the DEFECT item. Our analysis supports the conclusion that the NOMORE item is strongly affected by the context in which it was presented. Note, however, that the difference in the NOMORE and DEFECT items may reflect both context eefects as well as real change. In the present study we are not able to differentiate these two effects. Therefore, in the 1988 AAS we restored the 1984 order so that we have an ABA-quasi experimental design across the three time periods. The results of the 1988 AAS should help us to decide whether the differences in levels of endorsement of the items DEFECT and NOMORE between 1984 and 1987, represent

change over time, context effects, or a combination. If there are context effects, the proportion who endorse NOMORE will decrease drastically and the proportion who endorse DEFECT will increase. If the change is temporal, the increase in NOMORE will either continue to increase or level off, and the decrease in DEFECT will continue or level off.

Although the order of presentation may affect the proportion of respondents who endorse a particular item, our further analysis with the Mokken method revealed that the scalability of the six items remain the same. More specifically, the item difficulties may differ due to the the item manipulation but the relative ordinal positons of the items remained stable for both surveys. Consequently, we may conclude that the unidimensionality of the six abortion items is unaffected by the change in item difficulties, regardless of whether the change is due to context effects or "real" change over time. Thus, despite there is a change in item difficulties across the two years, we are still able to measure the change in general attitudes toward abortion.

References

Arney, W.R., and W.H. Trescher

1976 "Trends in attitudes toward abortion." Family Planning Perspectives 8: 117-124.

Barnett, S.H., and R.J. Harris

1982 "Recent changes in predictors of abortion attitudes." Sociology and Social Research 66: 320-324.

Bradburn, N.M., and S. Sudman

1979 Improving Interview Method and Questionnaire. San Francisco: Josey Bass.

Clogg, C.C., and D.O. Sawyer

1981 "A comparison of Alternative models for analyzing the scalability of

response patterns." Pp. 240-280 in S. Leinhardt (ed.), Sociological Methodology. London: Josey-Bass Publishers.

Combs, M., and S. Welch

1982 "Blacks, whites, and attitudes toward abortion." Public Opinion Quarterly 46: 510-520.

Duncan, O., and H. Schuman

1977 "An experiment on order and wording of questions." Unpublished manuscript, Department of Sociology, University of Arizona.

Ebaugh, H.R., and C.A. Haney

1980 "Shifts in abortion attitudes: 1972-1980." Journal of Marriage and the Family 42: 491-499.

Gillespie, M.W., Ten Vergert, E.M., and J. Kingma

1987a "Using Mokken scale analysis to develop unidimensional scales: Do the six abortion items in the NORC GSS form one or two scales". Quality and Quantity: International Journal of Methodology" 21: 1-15.

Gillespie, M.W., Ten Vergert, E.M., and J. Kingma

1987b "Using Mokken methods to develop robust cross-national scales: American and West German attitudes toward abortion". Social Indicators Research 19: 74-95.

Gillespie, M.W., Ten Vergert, E.M., and J. Kingma

1988 "Secular Trends in abortion attitudes; 1975-1980-1985". Journal of Psychology in press.

Granberg, D.

1978 "Pro-life or reflection of conservative ideology? An analysis of opposition to legalized abortion". Sociology and Social Research 62: 414-429.

Granberg, D., and B. Granberg

1980 "Abortion attitudes, 1965-1980: Trends and determinants". Family Planning

and Perspective 12: 250-261.

Guttman, L.

1950 "The basis for scalogram analysis," Pp. 60-90 in S.A. Stouffer, L. Guttman, P.F. Lazarsfeld, S.A. Star, and J.A. Clausen (eds.). Measurement and prediction. Studies in social psychology in World war II. Vol. IV. Princeton (N.J.): Princeton University Press.

Hall, E.J.. and M.M. Ferree

1986 "Race differences in abortion attitudes." Public Opinion Quarterly 50: 193-207.

Hartnagel, T.F., Creechan, J.J., and R.A. Silverman

1985 "Public opinion and the legalization of abortion." Canadian Review of Sociology and Anthropology 22: 411-430.

Hattie, J.

1985 "Methodological review: Assessing the unidimensionality and tests of items." Applied Psychological Measurement 9: 139-164.

Jones, E.F., and C.F. Westhoff

1973 "Changes in attitudes toward abortion: with emphasis upon the national fertility study data" Pp. 468-481 in H.J. Osofsky and D.J. Osofsky (eds.), The abortion experience: Psychological and medical impact. New York: Harper and Row.

Kingma, J., and J. Reuvekamp

1986a "Mokken model: A Pascal program for nonparametric scaling." Educational and Psychological Measurement 46: 667-677.

Kingma, J., and J. Reuvekamp

1986b "Mokken test for the robustness of nonparametric scaling." Educational and Psychological Measurement 46: 679-685.

Kinzel, C.

1984 Edmonton Area Series Report No. 33. Edmonton Area Study Sampling Report 1984. Population Research Laboratory, Department of Sociology, University of Alberta.

Kinzel, C.

1987 Edmonton Area Series Report No. 51. All Alberta Study Sampling Report 1987. Population Research Laboratory, Department of Sociology, University of Alberta.

Kornhauser, A., and P. Sheatsley

1976 "Questionnaire construction and interview procedure" in C. Sellitz, L. Wrightsman, and S. Cook. Research Methods in Social Relations. New York: Holt, Rinehart, & Winston.

Kraut, A.I., Wolfson, A.D., and A. Rothenberg

1975 "Some effects of position on opinion survey items." Journal of Applied Psychology 60: 774-776.

Loevinger, J.

1948 "The technique for homogeneous tests." Psychological Bulletin 45: 507-529.

McFarland, S.M.

1981 "Effects of Question Order on Survey Responses." Public Opinion Quarterly 45: 208-215.

Mokken, R.J.

1971 A theory and procedure of scale analysis. The Hague: Mouton.

Molenaar, I.W.

1983 "Rasch, Mokken en schoolbeleving" (Rasch, Mokken and school experience) Pp. 195-213 in S. Lindenberg and F.N. Stokman (eds.). Modellen in de Sociologie (Models in sociology). Deventer: Van Loghum Slaterus.

Niemoller, B., Verschuur, W., and F.R. Stokman

1980 Stochastic cumulative scaling: STAP User's Manual. Vol. 4. Amsterdam: University of Amsterdam.

Nunally, J.M.

1978 Psychometric Theory. New York: McGraw-Hill Book Company.

Schuman, H., Presser, S., and J. Ludwig

1981 "Context effects on survey responses to questions about abortion." Public Opinion Quarterly 45: 216-223.

Sigelman, L.

1981 "Question-order effects on presidential popularity." Public Opinion Quarterly 45: 199-207.

SPSS User's Guide (1983). New York: McGraw Hill Book Company.

Stinchombe, A.L.

1983 "Linearity in log-linear analysis" in S. Leinhardt (ed.). Sociological Methodology 1983-1984. London: Jossey-Bass.

Ten Berge, J.

1972 "Difficulty factors, distribution effects, and the least squares data matrix solution." Educational and Psychological Measurement 32: 911-920.

Ten Vergert, E.M.

1988 A survey of scaling methods for the measurement of attitudes: A comparison of American and West German attitudes towards abortion. Unpublished doctoral dissertation.

APPENDIX A

The Assumption of Double Monotony

The Mokken model treats the attitude as a single latent trait on which the person's location is represented by the parameter θ and the items's location is represented by parameter δ. Given a reasonably unidimensional set of items, that is, one dominated by the latent trait being measured, the person parameter (θ) can be estimated by the number of items to which the person responds positively, and the item parameter (δ) can be estimated by the proportion of people who respond respond positively. The former is referred to as the person's scale score and the latter as the item difficulty.

The relationship between the person parameter θ and the probability of a positive response on a particular item is described by means of an item trace line. The Mokken model makes no assumption about the functional form of the item characteristic curve or ICC. Instead, the only constraint the Mokken model puts on the ICC's is referred to as the assumption of double monotony. The first requirement of this assumption, the monotone homogeneity, is that for any item in a Mokken scale, the probability of a positive response increases as θ increases. To put this more formally, for any two persons i and j, where $\theta_i < \theta_j$ the probability of a positive response on any item in the scale is less for person i.

The second requirement of the assumption of double monotony, the monotonicity in item difficulties, is that for any value of θ the probability of a positive response decreases with the difficulty of the item. This means that the order of item difficulties remains invariant over the values of θ, or put graphically, the ICC's do not intersect. Given this assumption, it becomes possible to define unambiguously the difficulty of an item with a probability of 0.5 (Mokken, 1971).

Figure A1 graphically illustrates the proporty of a Mokken scale. It contains the ICC's of four different items. Item 4 is the easiest, followed by items 3, 2, and 1 in order of increasing difficulty. The items satisfy the assumption of double monotony; the ICC for each item increases with θ, and none of the ICC's intersect. Note that the value of δ for each item is found by drawing a line from the ICC to the θ axis at the point on the ICC where the probability of a positive response is 0.5. Also note that the ordinal scale scores are defined in terms of the (unobserved) values of δ. Finally, note that, within the constraints of double monotony, the functional form of the ICC's differ.

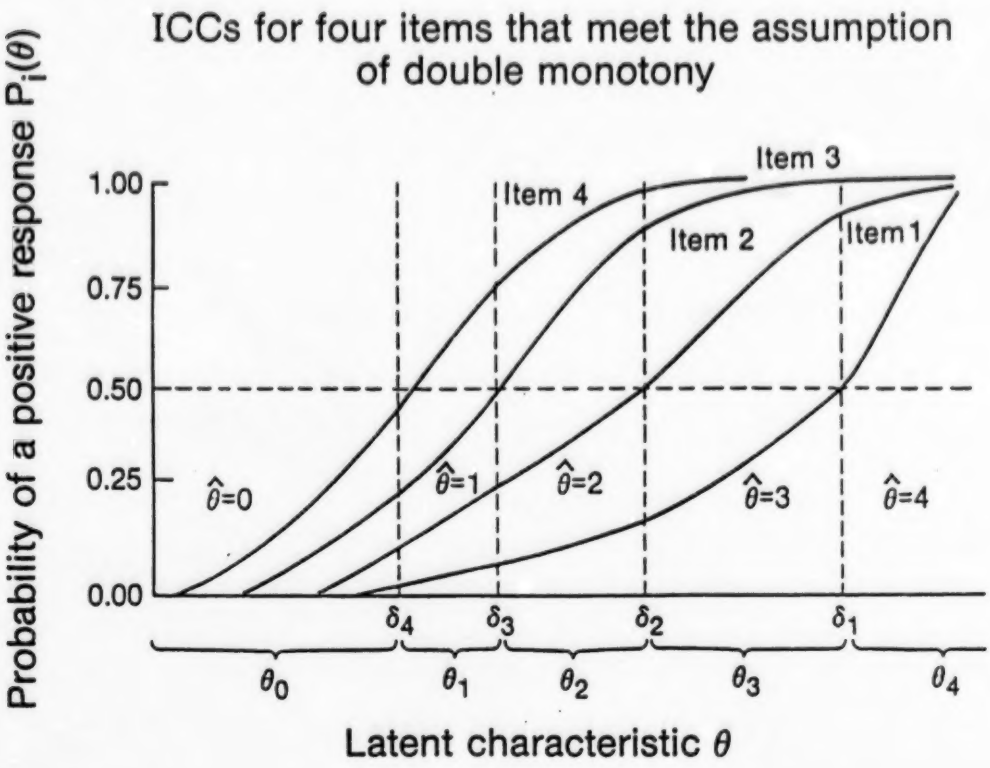
The value of double monotony as a criterion lies in the balance it strikes between flexibility and rigor. On the one hand, by permitting different functional forms of the ICC, it avoids the problems discussed by Hattie (1985), Ten Berge (1972), and Stinchombe (1983). On the other hand, it implies a sufficient constraint on the functional form. The test of the assumption of double monotony implies both testing the monotone homogeneity of the items and testing the monotonicity in the item difficulties.

Testing the Monotone Homogeneity of a Set of Items.

Mokken (1971) developed three related coefficients of scalability to test the homogeneity of a set of items. The first, H_{ij}, measures the homogeneity or association between each pair of items. The second, H_i, measures the homogeneity of a particular item with respect to all other items and is obtained by aggregating across the psychometric equivalence coefficients for the relevant item pairs. The third, H, measures the homogeneity of the scale as a whole by aggregating across the coefficients for the individual items.

Due to Loevinger (1948), the coefficient for the homogeneity of an item pair essentially measures the association in the two by two table that is

Figure A1: An example of four double monotone tracelines.



obtained by cross classifying the two items. Table A1 illustrates such a table. In contructing this table, we assumed that item $\underline{i}$ is more difficult than item $\underline{j}$, and let item $\underline{i}$ define the rows and item $\underline{j}$ define the columns.

Under Guttman's deterministic model we would expect the top right-hand cell, the "error cell", to be empty, i.e., $\underline{f}(1,0)=0$. Under the model of statistical independence (or no association between $\underline{i}$ and $\underline{j}$), we would expect the frequency of the error cell equal the product of the marginal frequencies divided by the sample size, i.e., $\underline{e}(1,0)$ would equal $\underline{f}(1,.)\underline{f}(.,0)/\underline{f}(.,.)$. Given in (1), H_{ij}, the index of item pair homogeneity measures the proportional difference between cell frequency of the error cell expected under independence and the actual cell frequency.

$$H_{ij} = \frac{[\underline{e}(1,0)-\underline{f}(1,0)]}{\underline{e}(1,0)} \qquad (1)$$

Where $\underline{e}(1,0)=\underline{f}(1,.)\underline{f}(.,0)/\underline{f}(.,.)$.

Readers familiar with the convention of using the letters $\underline{a}$, $\underline{b}$, $\underline{c}$, and $\underline{d}$ to represent the cell frequencies of a 2x2 table may find the following formula for H_{ij} more convenient.

$$H_{ij} = (\underline{ad}-\underline{bc})/(\underline{a}+\underline{b})(\underline{b}+\underline{d}). \qquad (2)$$

They also may notice the similarity between this index and other measures of association for 2x2 tables. When the items are independent, H_{ij} will be zero; when the error cell is empty, H_{ij} will equal unity.

The coefficient of item homogeneity, H_i, is given in (3). It simply aggregates the observed and expected frequencies used to calculate H_{ij} for all item pairs from a set of $\underline{k}$ items that contain item $\underline{i}$.

Table A1: The Cross-tabulation of Two Items

Response to item i	Response to item j		Row Total
	1	0	
1	$f(1,1)$	$f(1,0)$	$f(1,.)$
0	$f(0,1)$	$f(0,0)$	$f(0,.)$
Column Total	$f(.,1)$	$f(.,0)$	$f(.,.)$

Note: Item i is assumed to be more difficult than item j. "1" denotes a positive response; "0" denotes a negative response.

$$H_i = \frac{\sum_{j-1}^{k} e_{ij} - \sum_{j-1}^{k} f_{ij}}{\sum_{j-1}^{k} e_{ij}} \qquad (3)$$

where $i \neq j$.

The coefficient of scale homogeneity, H is given in (4). It aggregates the observed and expected frequencies used to calculate H_{ij} for all item pairs.

$$H = \frac{\sum_{i-1}^{k-1} \sum_{j-i+1}^{k} e_{ij} - \sum_{i-1}^{k-1} \sum_{j-i+1}^{k} f_{ij}}{\sum_{i-1}^{k-1} \sum_{j-i+1}^{k} e_{ij}} \qquad (4)$$

where $i-j+1$.

The coefficients of scale and item homogeneity allow the researcher to judge the scale as a whole and the scalability of individual items. Computer programs developed by Niemoller (1980) and Kingma and Reuvekamp (1986a, 1986b) contain such a search procedure.

Testing the Monotonicity in the Item Difficulties

The three coefficients of scalability are used to test the monotone homogeneity of the items. This test is a necessary, but not sufficient, condition to test the assumption of double monony underlying the Mokken scaling model. The test of both the monotonicity of item difficulties and the

test for monotone homogeneity of the items provide a sufficient condition to test this assumption.

The test of the monotonicity of item difficulties involves an inspection of two matrices, the $\underline{P}$ and $\underline{P}_0$ matrices which contain the probabilities of two positive and two negative responses, respectively, to all possible pairs of items. This test is based on the assumption of local independence of item responses for a fixed person's parameter value. More specific, item responses are conditionally independent given the same value of θ, so that the conditional probability of joint responses (to items $\underline{i}$ and $\underline{j}$) for respondents with the same value of θ equals the product of their marginal probabilities of those responses (Mokken, 1971). Thus, when items 1, 2, and 3 represent decreasing levels of difficulty, the probability of a pair of positive responses will be the greatest for items 2 and 3, followed by 1 and 3, and 1 and 2. Similarly, the probability of a pair of negative responses will be greatest for 1 and 2, followed by 1 and 3, and 2 and 3.

The test of double monotony, then, involves an inspection of the $\underline{P}$ and $\underline{P}_0$ matrices. When the rows and columns are ordered from top to bottom and from left to right according to decreasing levels of item difficulty, the probability of a pair of positive responses should increase in the ($\underline{P}$-matrix), while the probability of a pair of negative responses should decrease in the ($\underline{P}_0$-matrix). Deviations from this pattern can be tested with the McNemar test (see Molenaar, 1983). Items which do not fit this pattern are removed from the scale although allowances are made for random error.

APPENDIX B

Table 1B: The $\underline{P}$ and $\underline{P}_0$ matrix for the Six Abortion Items, 1984

Item	S	P	N	R	D	H
			Item			
			$\underline{P}$ matrix			
S	--	.33	.33	.39	.39	.39
P	.33	--	.37	.41	.41	.42
N	.33	.37	--	.42	.42	.42
R	.39	.41	.42	--	.78	.81
D	.39	.41	.42	.78	--	.84
H	.39	.42	.42	.81	.84	--
			$\underline{P}_0$ matrix			
S	--	.51	.50	.16	.13	.09
P	.51	--	.51	.16	.13	.09
N	.50	.51	--	.15	.13	.09
R	.16	.16	.15	--	.09	.08
D	.13	.13	.13	.09	--	.08
H	.09	.09	.09	.08	.08	--

Note: The letters S, P, N, R, D, and H refer, respectively, to the six abortion items SINGLE, POOR, NOMORE, RAPE, DEFECT, and HEALTH.

Table 2B: $\underline{P}$ and $\underline{P}_0$ matrix for the Six Abortion Items, 1987.

	Item					
	S	P	N	R	D	H
			$\underline{P}$ matrix			
S	--	.36	.38	.40	.40	.40
P	.36	--	.41	.43	.43	.43
N	.38	.41	--	.54	.55	.55
R	.40	.43	.54	--	.84	.84
D	.40	.43	.55	.84	--	.93
H	.40	.43	.55	.84	.93	--
			$\underline{P}_0$ matrix			
S	--	.52	.41	.15	.06	.06
P	.52	--	.41	.14	.06	.06
N	.41	.41	--	.13	.06	.06
R	.15	.14	.13	--	.05	.05
D	.06	.06	.06	.05	--	.06
H	.06	.06	.06	.05	.06	--

Note: The letters S, P, N, R, D, and H refer, respectively, to the six abortion items SINGLE, POOR, NOMORE, RAPE, DEFECT and HEALTH.

EDMONTON AREA SERIES REPORTS

44. Quality of the Urban Environment as Perceived by Residents of Slow and Fast Growth Cities.
    Raymond F. Currie and Charlene Thacker. September 1985.

45. Moving to Get Ahead: Does it Pay?
    Leslie W. Kennedy and N. Mehra. October 1985.

46. Accident History, Perceived Risk of Personal Injury, and Job Mobility as Factors Influencing Occupational Accident Fatalism.
    W. Andrew Harrell. November 1985.

47. The Redistributive Effects of Lotteries in Western Canada.
    John R. Livernois. February 1986.

48. Public Explanations of Unemployment in a Canadian City.
    G.S. Lowe, H. Krahn, J. Tanner and T.F. Hartnagel. March 1986.

49. Effects of Unemployment on Psychological Well-being.
    Wei-Jung J. Yeung and P. Krishnan. June 1986

50. Edmonton Area Study Sampling Report 1986.
    Clifford Kinzel. June 1986.

51. 1987 All Alberta Study (including the Edmonton Area Study) Sampling Report.
    Cliff Kinzel. May 1987.

52. Perceived Quality of Life in a Centre: Periphery Framework.
    J.R. Oppong, R.G. Ironside and L.W. Kennedy. June 1987.

53. The Incidence of Wife Assault in Alberta.
    L.W. Kennedy and D.G. Dutton. July 1987.

54. Mixed Motivation for Migration in the Urban Prairies: A Comparative Approach.
    Raymond F. Currie and Shiva S. Halli. October 1987.

55. Drunk-Driving Intervention in an Urban Community: An Exploratory Analysis.
    Akin Adebayo. October 1987.

56. Migration to a Canadian Metropolitan City: Characteristics and Motives.
    Akin Adebayo. November 1987.

57. Factors Antecedent to Impaired Driving: A Research Note.
    Akin Adebayo. January 1988.

58. All Alberta Study 1988 (Including The Edmonton Area Study) Sampling Report.
    Cliff Kinzel and Akin Adebayo. May 1988.

59. Recent Trends in Public Support for Unions.
    Graham S. Lowe and Harvey Krahn. May 1988.